UNIVERSITY OF TARTU
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
Institute of Mathematical Statistics

Liivika Tee

Stochastic Reserving Methods in Non-Life Insurance

Master's Thesis

Supervisor: Meelis Käärik, Ph.D

TARTU 2013

Contents

Introduction

A non-life insurance company has to set up a fund to enable the company to meet and administer its contractual obligations to policyholders. The amount of future compensation of policyholders is usually referred to as the claims reserve for claim events that have already occurred. Obviously, the profitability and solvency is important for insurance company and solvency of a business is highly dependent upon the reserve level and the reserving philosophy. Therefore, reserving is a fundamental aspect of business management.

Loss events with the number and amount of claims are random. It is important to calculate the claims reserve carefully, since underestimation would lead to solvency problems and overestimation unnecessarily holds the excess capital instead of using it for other purposes. The claims estimation is one of the basic actuarial tasks in the insurance industry, because it gives the certainty to be solvent at any time moment in the future. On the basis of historical data the actuary can obtain predictions of the expected outstanding claims. Since the uncertainty of the actuary's best estimate can be quite high it is important that the actuary's best estimate is complemented by some measure of variability.

There is a variety of methods for the actuary to choose amongst for reserving purposes. One popular statistical method is the chain-ladder method, which originally was a deterministic method and is widely used due to its simplicity. The chain-ladder method gives a point estimate, but interest arises in developing estimates of the likely variability of the claims reserve. The usual statistical approach would start with specifying a model, finding an estimate of the outstanding claims under that model and then finding the precision of the estimate. Stochastic claims reserving starts with constructing a model and a method that produces the actuary's best estimate and then using this model for estimating the prediction error of the model. Moreover, there is a tendency to find a model under which the best

estimate is the one given by the chain-ladder.

The aim of the present thesis is to describe the classical basic chain-ladder method and several stochastic methods. The thesis is set out as follows. First section starts with the notation and basic results. It is followed by the overview and description of the chain-ladder technique. The section continues with the Mack's stochastic model, where the model assumptions and the results of calculating the variability are given.

Section 2 provides an introduction to stochastic models in the basis of generalized linear models (GLM). Discussion starts with the (over-dispersed) Poisson model and since there are several models linked to Poisson model, these models are examined as well. The stochastic models are introduced with the ideas of constructing the models and since the main focus is on estimating the likely variability of the estimate, the results of prediction errors are given.

In section 3, the models considered in the previous chapter, will be compared. As the Mack's distribution-free model and the Poisson model are considered as the chain-ladder "type" methods, it is important to point out the main differences of these models. The comparison leads to the known fact, that the distributin-free model of Mack is called as *the* stochastic model underlying the chain-ladder method. In addition, discussion about possible negative increments and how the proposed methods deal with them is provided.

The last section provides a practical reserving approach. The theoretical results considered in previous sections are implemented in a practical numerical problem, the reserve estimates and their mean square errors (and standard errors) of predictions are found for models of Mack, over-dispersed Poisson, log-normal and Gamma.

1 The basic chain-ladder method

The most widely used reserving method is the basic chain-ladder method, mainly because it is simple and distribution-free, i.e. it seems to work nearly without assumptions. The chain-ladder algorithm was developed as a deterministic algorithm and did not have any stochastic model underlying it (Mack, Venter, 1999). Therefore it is asking the impossible to find out what model underlies that algorithm by historical inquiry or strict logical deduction. However, an underlying stochastic model is required for estimating the prediction error of the algorithm.

The idea behind the chain-ladder method is comparatively simple. Method is based on the assumption that proportionate relationships between values in consecutive development years will repeat in the future, i.e. the columns in the run-off triangle are proportional and hence it is possible to obtain forecasts of ultimate claims based on the observed data, where 'ultimate' denotes the latest delay year so far observed, and does not include any tail factor.

1.1 Notations and basic results

The straightforward chain-ladder technique uses cumulative data, but it is not relevant in principle whether incremental or cumulative data are used when considering claims reserving in a stochastic context. Whatever data are available can be used with corresponding model and the same results are obtained. Thus, without loss of generality we may assume that the data what have been collected for $i = 1, 2, ..., n$ and $j = 1, 2, ..., n$, consist of a triangle of incremental claims:

$$\{C_{ij} : \ i = 1, ..., n; \ j = 1, ..., n - i + 1\},$$

where row index i refers to the year of origin and according to situation indicates accident year, reporting year or underwriting year. The column index j refers to the development year indicating the delay, more precisely loss disbursal, reporting

year or accident year.

Claims data are given as a *run-off triangle* as follows:

	Development period j				
Year of origin i	1	2	3	...	n
1	C_{11}	C_{12}	C_{13}	...	C_{1n}
2	C_{21}	C_{22}	...		
3	C_{31}	...			
...	...				
n	C_{n1}				

where C_{ij} denotes the amount of claims, occurred in year of origin i, to be paid in development year j.

We define the cumulative claim amounts with accident year index i reported up to, and including, delay index j by:

$$D_{ij} = \sum_{k=1}^{j} C_{ik},$$

so D_{ij} is the total claims amount of accident year i, $i = 1, \ldots, n$, either paid or incurred up to development year j, $j = 1, \ldots, n$ and we consider D_{ij} of which we have an observation if $i + j \leq n + 1$. It is assumed D_{in} to be the ultimate claim amount.

The development factors of the chain-ladder technique, known also as age-to-age factors or link ratios, are denoted by

$$\{\lambda_j : j = 1, \ldots, n-1\}.$$

The chain-ladder technique estimates the development factors as:

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j} D_{i,j+1}}{\sum_{i=1}^{n-j} D_{ij}}.$$

In order to produce forecasts of future values of cumulative claims, we need to apply development factors to the latest cumulative claims in each row $(D_{i,n-i+1})$:

$$\hat{D}_{i,n-i+2} = D_{i,n-i+1}\hat{\lambda}_{n-i+1}$$

$$\hat{D}_{ik} = \hat{D}_{i,k-1}\hat{\lambda}_{k-1}, \quad k = n-i+3, n-i+4, \ldots, n.$$

Remark 1.1. *It is the basic assumption of the chain-ladder method that the column vectors $\{D_j : j = 1, 2, \ldots, n\}$ are, up to random fluctuations, proportional to each other, i.e.*

$$D_{i,j+1} \cong D_{ij}\lambda_j.$$

Let $D_{ij} > 0$ be the cumulative claims amount of accident year i, $i = 1, \ldots, n$, after j years of development $j = 1, \ldots, n$. The amounts D_{ij} with $i + j = 1, \ldots, n+1$ are observable and the aim is to estimate the ultimate claims amount D_{in} and the outstanding claims reserve

$$R_i = D_{in} - D_{i,n+1-i} \tag{1.1}$$

for accident years $i = 2, \ldots, n$. The chain-ladder method estimates the ultimate claims amount D_{in} recursively by

$$\hat{D}_{in} = D_{i,n+1-i} \cdot \hat{\lambda}_{n+1-i} \cdot \ldots \cdot \hat{\lambda}_{n-1} \tag{1.2}$$

and the reserve R_i is estimated equivalently by

$$\hat{R}_i = D_{i,n+1-i}\left(\hat{\lambda}_{n+1-i} \cdot \ldots \cdot \hat{\lambda}_{n-1} - 1\right).$$

Corollary 1.2. *Let $C = \{D_{ij} \mid i + j = 2, \ldots, n+1\}$, then the estimator $\hat{D}_{in}$ in (1.2) has the same form as $E(D_{in} \mid C)$, which is the best forecast of D_{in} based on the observation C.*

1.2 Mack's stochastic model

There are a number of reserving methods which have proved useful in practice, one of which is extensively used and is known as the chain-ladder technique (Verrall, 1994). The simplest assumption of this method is that payments will emerge in a similar way in each accident year. The proportionate increases in the known cumulative payments from one development year to the next can then be used to calculate the expected cumulative payments for future development years.

The chain-ladder method gives us a point estimate, but interest arises in developing estimates of the likely variability in the outcome. Therefore, Mack (1993) has developed a distribution-free method in order to estimate the prediction error of chain-ladder reserve estimates and in further section it will be shown, that Mack's model qualifies as *the* stochastic model underlying the chain-ladder algorithm.

Mack (1993) has introduced following basic chain-ladder *assumptions*:
(A1) there are development factors $\lambda_1, \lambda_2, \ldots, \lambda_{n-1} > 0$ with

$$E(D_{i,j+1} \mid D_{i1}, \ldots, D_{ij}) = D_{ij}\lambda_j,\ i = 1, \ldots, n,\ j = 1, \ldots, n-1;$$

(A2) the variables D_{ij} of different accident years, i.e.

$$\{D_{i1}, \ldots, D_{in}\}, \{D_{k1}, \ldots, D_{kn}\},\ i \neq k,\ \text{are independent};$$

(A3) with unknown proportionality constants σ_j^2, $j = 1, \ldots, n-1$,

$$Var(D_{i,j+1} \mid D_{i1}, \ldots, D_{ij}) = D_{ij}\sigma_j^2.$$

Remark 1.3 (Mack, 1993). *The assumption **(A2)** is made according to the fact that chain-ladder algorithm does not take into account any dependencies between accident years. However, in practice, the independence of the accident years can be distorted by certain calendar year effects like major changes in claims handling or in case reserving.*

Remark 1.4. *Failure of independency assumption **(A2)** usually requires a different approach. If the assumpiton **(A3)** is violated, some other variant of the chain-ladder algorithm might still be optimal.*

The following theorem, stated by Mack (1993), shows that under the results (1.2) and Corollary 2, where $\lambda_1, \ldots, \lambda_{n-1}$ is estimated by $\hat{\lambda}_1, \ldots, \hat{\lambda}_{n-1}$ is indeed a reasonable procedure.

Theorem 1.5. *Under the assumptions **(A1)** and **(A2)** the estimators $\hat{\lambda}_j$, $j = 1, \ldots, n-1$, are unbiased and uncorrelated.*

The unbiasedness of the λ_j is easy to see, but the uncorrelatedness of the $\hat{\lambda}_j$ might be surprising because $\hat{\lambda}_{j-1}$ and $\hat{\lambda}_j$ depend on the same data $D_{1j}, \ldots, D_{n-j,j}$.

The previous result easily extends to random products of pairwise different $\hat{\lambda}_j$, i.e. we have

$$E(\hat{\lambda}_{n+1-i} \cdot \ldots \cdot \hat{\lambda}_{n-1}) = \lambda_{n+1-i} \cdot \ldots \cdot \lambda_{n-1},$$

which shows that $\hat{D}_{in} = D_{i,n+1-i}\hat{\lambda}_{n+1-i} \cdot \ldots \cdot \hat{\lambda}_{n-1}$ is an unbiased estimator of $E(D_{in}|C) = D_{i,n+1-i}\lambda_{n+1-i} \cdot \ldots \cdot \lambda_{n-1}$. Similarly, the reserve estimator

$$\hat{R}_i = \hat{D}_{in} - D_{i,n+1-i} = \sum_{i=1}^{n} D_{i,n-i}(\hat{\lambda}_{n-i} \cdot \hat{\lambda}_{n-i+1} \cdot \ldots \cdot \hat{\lambda}_{n-1} - 1) \qquad (1.3)$$

is an unbiased estimator of the true reserve $R_i = D_{in} - D_{i,n+1-i}$.

1.3 The results of calculating mean squared error and standard error

Claims reserving is a predictive process: given the data, we try to predict future claims (England,Verrall, 2002). From a statistical viewpoint, given a point estimate, the natural next step is to develop estimates of the likely variability in the

outcome so that assessments can be made, for instance, of whether extra reserves should be held for prudence, over and above the predicted values.

The measure of variability commonly used is the *prediction error*. In this context we use the expected value as the prediction. The method developed by Mack (1993) for estimating the prediciton error is distribution-free and in this respect, certain unbiased estimators for the mean squared errors (MSE) are found.

The MSE of the (unbiased) prediction $\hat{D}_{ij}$ for each future observation D_{ij}, $i+j = n+2, \ldots, 2n$, can approximately be decomposed into two parts, process variance and estimation variance:

$$E[(D_{ij} - \hat{D}_{ij})^2] = E[((D_{ij} - E[D_{ij}]) - (\hat{D}_{ij} - E[D_{ij}]))^2],$$

where plugging in $\hat{D}_{ij}$ instead of D_{ij} in the final expectation and expanding gives

$$\begin{aligned} E[(D_{ij} - \hat{D}_{ij})^2] &\approx E[(D_{ij} - E[D_{ij}])^2] - 2E[(D_{ij} - E[D_{ij}])(\hat{D}_{ij} - E[\hat{D}_{ij}])] \\ &+ E[(\hat{D}_{ij} - E[\hat{D}_{ij}])^2]. \end{aligned}$$

Assuming the future loss D_{ij} and its forecast $\hat{D}_{ij}$ computed from past losses to be independent, we get:

$$E[(D_{ij} - \hat{D}_{ij})^2] = E[(D_{ij} - E[D_{ij}])^2] + E[(\hat{D}_{ij} - E[\hat{D}_{ij}])^2],$$

hence the MSE is the sum of process variance and estimation variance. However, in the following we use the conditional mean squared error, since the unconditional MSE

$$E[(D_{ij} - \hat{D}_{ij})^2] = E[E((D_{ij} - \hat{D}_{ij})^2 |\, C)]$$

averages over all possible data C from the underlying distribution. In practice, there is more interest in the conditional MSE of the particular estimated amount $\hat{D}_{in}$ based on the specific data set C observed and therefore have to use $E((\hat{D}_{in} - D_{in})^2 |\, C)$ which just gives us the average deviation between $\hat{D}_{in}$ and D_{in} due to future randomness only (Mack, 1993).

The (conditional) mean squared error $MSE(\hat{D}_{in}|\, C)$ of the estimator $\hat{D}_{in}$ of D_{in} is defined as

$$MSE(\hat{D}_{in}|\, C) = E[(D_{in} - \hat{D}_{in})^2|\, C].$$

Using calculations similar to unconditional case, we can write this as:

$$\begin{aligned} MSE(\hat{D}_{in}|\, C) &= E[(D_{in} - E[D_{in}] + E[D_{in}] - \hat{D}_{in})^2|\, C] \\ &= Var[D_{in}|\, C] + E[E[D_{in}|\, C] - \hat{D}_{in}]^2 \\ &= Var[D_{in}|\, C] + (E[D_{in}|\, C] - \hat{D}_{in})^2, \end{aligned}$$

where $C = \{D_{ij} \mid i + j = 2, \ldots n + 1\}$ is the set of all data observed so far. Therefore, if an estimate is unbiased, its MSE is equal to its variance, which makes MSE a convenient choice to model the variability of the predictions.

Using the equations from results (1.1) and (1.3), we see that

$$\begin{aligned} MSE(\hat{R}_i) &= E[(R_i)^2 - \hat{R}_i|\, C] \\ &= E[(D_{in} - D_{i,n+1-i})^2 - \hat{D}_{in} + D_{i,n+1-i}|\, C] \\ &= E[(D_{in})^2 - \hat{D}_{in}|\, C] \\ &= MSE(\hat{D}_{in}), \end{aligned}$$

i.e. mean squared error of the reserve estimator $\hat{R}_i$ equals to mean squared error of ultimate claims amount estimator $\hat{D}_{in}$.

Next, because of the variance assumption (see the assumption **(A3)**), which is implicitly underlying the chain-ladder method, we will need an estimator for σ_j^2. Likewise in case of $\hat{\lambda}_j$, it is shown (Mack, 1993) that

$$\hat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=1}^{n-j} D_{ij} \left(\frac{D_{i,j+1}}{D_{ij}} - \hat{\lambda}_j \right)^2, \; j = 1, \ldots, n-2,$$

is an unbiased estimator of σ_j^2, $j = 1, \ldots, n-2$. We need an estimator for σ_{n-1} as well. If $\hat{\lambda}_{n-1} = 1$ and if the claims development is assumed to be finished after

$n-1$ years, we can put $\hat{\sigma}_{n-1} = 0$, otherwise it is required to use the estimator

$$\hat{\sigma}^2_{n-1} = \min(\frac{\hat{\sigma}^4_{n-2}}{\hat{\sigma}^2_{n-3}}, \min(\hat{\sigma}^2_{n-3}, \hat{\sigma}^2_{n-2})).$$

Now we can state the main result of the chain-ladder model with the following theorem, formulated by Mack.

Theorem 1.6. *Under the assumptions **(A1)**, **(A2)** and **(A3)** the mean squared error $MSE(\hat{R}_i)$ can be estimated by*

$$\widehat{MSE(\hat{R}_i)} = \hat{D}^2_{in} \sum_{j=n+1-i}^{n-1} \frac{\hat{\sigma}^2_j}{\hat{\lambda}^2_j} \left(\frac{1}{\hat{D}_{ij}} + \frac{1}{\sum_{k=1}^{n-j} D_{kj}} \right),$$

where $\hat{D}_{ij} = D_{i,n+1-i}\hat{\lambda}_{n+1-i} \cdot \ldots \cdot \hat{\lambda}_{j-1}$, $j > n+1-i$, are the estimated values of the future D_{ij} and $\hat{D}_{i,n+1-i} = D_{i,n+1-i}$.

Recall that the MSE equals to squared standard error plus squared bias. In our context we have unbiased estimate of R_i (see (1.3)), therefore the standard error $se(\hat{R}_i)$ is equal to the root mean square error of $\hat{R}_i$ (Holton, 2013):

$$se(\hat{R}_i) = \sqrt{MSE(\hat{R}_i)}.$$

If the assumptions **(A1)-(A3)** hold, the overall reserve equals to $R = \sum_{i=2}^{n} \sum_{j=n+1-i}^{n} D_{ij}$ and often the standard error of the overall reserve estimate $\hat{R} = \hat{R}_2 + \ldots + \hat{R}_n$ is of interest, too.

Corollary 1.7 (**Mack, 1993).** *Given the Theorem 1.6 the mean squared error of the overall reserve estimate $\hat{R} = \hat{R}_2 + \ldots + \hat{R}_n$ can be estimated by*

$$\widehat{MSE(\hat{R})} = \sum_{i=2}^{n} \left\{ (se(\hat{R}_i))^2 + \hat{D}_{in} \left(\sum_{k=i+1}^{n} \hat{D}_{kn} \right) \sum_{j=n+1-i}^{n-1} \frac{\frac{2\hat{\sigma}^2_j}{\hat{\lambda}^2_j}}{\sum_{m=1}^{n-j} D_{mj}} \right\}. \quad (1.4)$$

In conclusion, we introduced the chain-ladder algorithm to obtain forecasts of the reserves and formulas for MSEs of the claim reserve amounts that only use data from the chain-ladder triangle are given and these formulas allow us to calculate the variability of the proposed predictions. Forecasting outstanding claims and setting up suitable reserves to meet these claims is an important part of an actuary's tasks and in insurance industries in general. Thus, subjectivity and experience of an actuary are always suggested in estimating reserves, since not all the assumptions required in mathematical formulas are fulfilled in practice.

2 Generalized Linear Models

Generalized linear modeling is a methodology for modeling relationships between variables. It generalizes the classical normal linear model, by relaxing some of its restrictive assumptions, and provides methods for the analysis of non-normal data. The exponential family of distributions is one of the key constructs in generalized linear modeling (de Jong, Heller, 2008). Generalized Linear Model (GLM) is important in the analysis of insurance data, because with insurance data the assumptions of the normal model are often not applicable.

The traditional chain-ladder method is merely an algorithm, a deterministic method giving only point estimates. To estimate the variability of predicted claims reserve, a stochastic model is required. In this section we consider certain stochastic reserve estimation methods that also allow us to estimate the likely variability in the outcome.

Several often used and traditional actuarial methods to complete a run-off triangle can be described by one generalized linear model. Let the random variables C_{ij} for $i = 1, \ldots, n$ and $j = 1, \ldots, n$ denote the claim figure for year of origin i and year of development j, as before. We consider the following multiplicative model, with a parameter for each row i, each column j and each diagonal $k = i + j - 1$:

$$C_{ij} \approx \alpha_i \cdot \beta_j \cdot \gamma_k, \tag{2.1}$$

where parameter α_i describes the effect of year of origin i, parameter β_j corresponds to development year j and γ_k describes the effect of calendar year $k = i + j - 1$. (Kaas et al, 2008)

In the terminology of GLM, to linearise the multiplicative model (2.1) we choose the logarithm as a link function *(log-link)*. Thus, the expected value of C_{ij} can be

given as

$$\begin{aligned} EC_{ij} &= \alpha_i \cdot \beta_j \cdot \gamma_k \\ &= \exp(\ln \alpha_i + \ln \beta_j + \ln \gamma_k), \end{aligned}$$

or, equivalently,

$$\ln EC_{ij} = \ln \alpha_i + \ln \beta_j + \ln \gamma_k.$$

Having found the estimates of the parameters, it is easy to complete the run-off triangle, simply by taking

$$\hat{C}_{ij} := \hat{\alpha}_i \cdot \hat{\beta}_j \cdot \hat{\gamma}_k. \tag{2.2}$$

Turns out that simple model considered in this section allows generate quite a few reserving techniques, depending on the assumptions set on distribution of the C_{ij}. We proceed this section with deriving the following methods from model (2.1).

2.1 The Poisson model

When finding a stochastic model that reproduces chain-ladder estimates, some assumptions must be made about the insurance claims. It is possible either to specify the distribution of the insurance claims, or merely state the two first moments (England, Verrall, 2002).

There is a wide range of stochastic reserving models and they can be divided as chain ladder "type" and as extensions to the chain ladder. The chain-ladder "type" models may reproduce the chain ladder results exactly or can have a similar structure to chain-ladder without giving the exactly the same results. In this section we consider the Poisson model which reproduces reserve estimates given by the chain-ladder technique.

Remark 2.1 (England, Verrall, 2002). *Renshaw & Verrall (1998) were not the first to notice the link between the chain-ladder technique and the Poisson distribution, but were the first to implement the model using standard methodology in statistical modelling.*

Already in 1975 a stochastic model corresponding to Poisson model, which leads to the chain-ladder technique, was discovered. This model works on the incremental amounts

$$C_{ij} = D_{ij}, \text{ if } j = 1,$$

$$C_{ij} = D_{ij} - D_{i,j-1}, \text{ if } j > 1.$$

The model makes the following *assumptions*:

(P1) $E(C_{ij}) = x_i y_j$ with unknown parameters x_i and y_j.

(P2) Each incremental amount C_{ij} has a Poisson distribution.

(P3) All incremental amounts C_{ij} are independent.

Here x_i is the expected ultimate claims amount (up to the latest development year so far observed) and y_j is the proportion of ultimate claims to emerge in each development year with restriction $\sum_{k=1}^{n} y_k = 1$. The restriction immediately follows from the fact that y_j is interpreted as the proportion of claims reported in development year j. Obviously, the aggregate proportion over all periods has to be 1.

We estimate the unknown parameters x_i and y_j from the triangle of known data (notation Δ is used for that) with the maximum likelihood method. The estimation procedure and results are given with the following lemma.

Lemma 2.2. *Assume that all C_{ij} are independent with a Poisson distribution and $E(C_{ij}) = x_i y_j$ holds. Then the maximum likelihood estimators for x_i and y_j are given by:*

$$x_i = \frac{\sum_{j \in \Delta_i} C_{ij}}{\sum_{j \in \Delta_i} y_j}$$

$$y_j = \frac{\sum_{i\in\Delta_j} C_{ij}}{\sum_{i\in\Delta_j} x_i}.$$

Proof. We derive the maximum likelihood estimates for the unknown parameters x_i and y_j with the likelihood function

$$L = \prod_{i,j\in\Delta} \frac{(x_i y_j)^{C_{ij}}}{C_{ij}!} \exp(-x_i y_j).$$

Therefore the loglikelihood function is

$$\ell = \ln(L) = -\sum_{i,j\in\Delta} x_i y_j + \sum_{i,j\in\Delta} C_{ij} \ln(x_i y_j) - \sum_{i,j\in\Delta} \ln(C_{ij}!),$$

where the summation is for all i, j where C_{ij} is known. The maximum likelihood estimator are those values x_i, y_j which maximize L or equivalently $\ln(L)$. They are given by the equations

$$0 = \frac{\partial \ell}{\partial x_i} = -\sum_{j\in\Delta_i} y_j + \sum_{j\in\Delta_i} C_{ij} \frac{1}{x_i}$$

$$0 = \frac{\partial \ell}{\partial y_j} = -\sum_{i\in\Delta_j} x_i + \sum_{i\in\Delta_j} C_{ij} \frac{1}{y_j},$$

thus the likelihood estimator for x_i and y_j are given, respectively, by

$$x_i = \frac{\sum_{j\in\Delta_i} C_{ij}}{\sum_{j\in\Delta_i} y_j}$$

$$y_j = \frac{\sum_{i\in\Delta_j} C_{ij}}{\sum_{i\in\Delta_j} x_i}.$$

The lemma is proved. ■

Thus, the proportion factors y_j express the ratio of the sum of observed incremental values for certain development year j with respect to certain ultimate claims, i.e. y_i denotes the proportion of claims reported in development year j. The parameters x_i refer to the ratio of the sum of observed incremental values for certain origin year i with respect to corresponding proportion factors, i.e. if the

incremental claim amounts and respective proportions factors are known, it is simple to derive the corresponding ultimate claim x_i for origin year i. One can note the principal similarities with chain-ladder technique, where development factors are also outcomes of certain ratios.

The mean given as $E(C_{ij}) = x_i y_j$ in assumption **(P1)** has a multiplicative structure, i.e. it is the product of the row effect and the column effect. Both the row and the column effect have specific interpretations and it is sometimes useful to preserve the model in this form. Nevertheless, for estimation purposes, sometimes it is better to reparameterise the model so that the mean has a linear form. The Poisson model can be cast into the form of a GLM and to linearise the multiplicative model we need to choose the logarithm as a link function so that

$$E(C_{ij}) = \exp(\ln x_i + \ln y_j),$$

or, equivalently,

$$\ln(E(C_{ij})) = \alpha_i + \beta_j, \tag{2.3}$$

where $\alpha_i = \ln(x_i)$ and $\beta_j = \ln(y_j)$ and structure of linear predictor (2.3) is still a chain-ladder type, because parameters for each row i and each column j are given. Hence, the structure (2.3) is defined as a generalised linear model, in which the incremental values C_{ij} are modelled as Poisson random variables with a logarithmic link function and linear predictor

$$\eta_{ij} = c + \alpha_i + \beta_j. \tag{2.4}$$

In any case, a constraint

$$\alpha_1 = \beta_1 = 0$$

is needed to estimate the remaining model parameters c, α_i, β_j and to avoid over-parametrization. Considering a single incremental payment C_{ij} with origin year i and claim payments in development year j (yet to be observed), we obtain the

estimates of future payments from the parameter estimates by inserting them into equation (2.3) and exponentiating, resulting as

$$\hat{C}_{ij} = \hat{x}_i \hat{y}_j = \exp(\hat{\eta}_{ij}). \tag{2.5}$$

Given the equation (2.5), the reserve estimates for origin year and overall estimates can be easily derived by summation:

$$\hat{R}_i = \hat{x}_i \hat{y}_{n+2-i} + \ldots + \hat{x}_i \hat{y}_n \tag{2.6}$$

From the assumptions (**P1**) - (**P3**), the maximum likelihood estimator (2.6) of the claims reserve for origin year i, $R_i = C_{n+2-i} + \ldots + C_{in} = D_{in} - D_{i,n+1-i}$, gives the same prediction $\hat{D}_{in} = D_{i,n+1-i} + \hat{R}_i$ as the chain-ladder method. According to the assumption **(P3)**, $D_{i,n+1-i} + \hat{R}_i$ is an estimator of the conditional expectation $E(D_{in} \mid D_{i1}, \ldots, D_{i,n+1-i})$ and assumption **(P2)** constrains all incremental amounts C_{ij} to be non-negative integers.

In the Poisson model for loss reserving it is assumed that the incremental claims are independent and Poisson distributed with expectations being the product of two factors, depending on the occurrence year and the development year, respectively. It is well-known that maximum-likelihood estimation in the Poisson model yields the chain-ladder estimators of the expected ultimate aggregate claims (Schmidt, 2002). Moreover, Renshaw & Verrall (1998) pointed out that this is also true for overdispersed Poisson models.

We recall, that the only distributional assumptions used in GLMs are the functional relationship between variance and mean and the fact, that distribution belongs to the exponential family. In case of Poisson, the mentioned relationship is $Var(C_{ij}) = E(C_{ij})$ and it can be generalized to $Var(C_{ij}) = \phi E(C_{ij})$ without any change in form and solution of the likelihood equations (Mack, Venter, 1999). This kind of generalisation allows for more dispersion in the data. For the

solution of the likelihood equations it is not needed incremental values C_{ij} to be non-negative or integers and this leads to an over-dispersed Poisson model and to quasi-likelihood equations, since the range of the underlying distribution is not important anymore.

2.2 The over-dispersed Poisson model

The model we consider in this section is based on the Poisson distribution. The specification of the Poisson modelling distribution does not mean that the model can only be applied to data which are positive integers; it is easy to write down a quasi-likelihood which has all the characteristics of a Poisson likelihood, without actually referring directly to the probability function for the Poisson random variable (Renshaw, Verrall, 1998). This means that the model can be applied to non-integer data, positive and negative.

The over-dispersed Poisson (ODP) model is different from the distribution-free chain-ladder model of Mack (1993), but both methods reproduce the historical chain-ladder estimator for the claims reserve and these models are the only ones known that lead to the same estimators for D_{in} as the chain-ladder algorithm. However, only the Mack's distribution-free model is close enough to the chain-ladder algorithm in enough aspects so it qualifies to be called *the* stochastic model underlying the chain-ladder algorithm, because the Poisson model deviates from the historical chain-ladder algorithm in several aspects that the Mack's distribution-free model does not (Mack, Venter, 1999).

The ODP distribution differs from the Poisson distribution in that the variance is not equal to the mean, but is *proportional* to the mean. It is shown (in Schmidt, 2002) that every ODP model can be transformed into the Poisson model by dividing all incremental claims by the parameter. The over-dispersed Poisson model assumes that the incremental claims C_{ij} are distributed as indepen-

dent over-dispersed Poisson random variables and the general form for the over-dispersed Poisson chain-ladder model can be given as follows:

$$E(C_{ij}) = x_i y_j,$$

$$Var(C_{ij}) = \phi x_i y_j,$$

where

$$\sum_{k=1}^{n} y_k = 1.$$

Over-dispersion is introduced through the parameter ϕ, which is unknown and estimated from the data.

Remark 2.3. *The parameter y_j appears in variance, so the restriction that y_j must be positive is automatically imposed. This leads to the limitation of the model that the sum of incremental claims in column j must be positive. Some negative incremental values are allowed, as long as any column sum is not negative.*

The over-dispersed Poisson model makes the following *assumptions* :

(ODP1) $E(C_{ij}) = x_i y_j$ with unknown parameters x_i, y_j.

(ODP2) The distribution of C_{ij} belongs to the exponential family with $Var(C_{ij}) = \phi x_i y_j$, where ϕ is an unknown parameter.

(ODP3) All C_{ij} are independent.

The resulting quasi-likelihood equations are

$$\sum_{j=1}^{n+1-i} x_i y_j = \sum_{j=1}^{n+1-i} C_{ij},\ i = 1, \ldots, n,$$

$$\sum_{i=1}^{n+1-j} x_i y_j = \sum_{i=1}^{n+1-j} C_{ij},\ j = 1, \ldots, n.$$

Mack (1991) has shown that these equations have the unique solution (if all $\hat{\lambda}_j$ are well defined and are not equal to zero, but without any restrictions on the row

sums or column sums over C_{ij}) :

$$\hat{x}_i\hat{y}_j = D_{i,n+1-i}\hat{\lambda}_{n+2-i} \cdot \ldots \cdot \hat{\lambda}_{j-1} \cdot (\hat{\lambda}_j - 1) \text{ for } j > n+1-i,$$

$$\hat{x}_i\hat{y}_j = D_{i,n+1-i}((\hat{\lambda}_{j+1} \cdot \ldots \cdot \hat{\lambda}_{n+1-i})^{-1} - (\hat{\lambda}_j \cdot \ldots \cdot \hat{\lambda}_{n+1-i})^{-1}) \text{ for } j \leq n+1-i,$$

with $\hat{\lambda}_j$ from the chain-ladder algorithm.
Because $(\hat{\lambda}_j - 1) + \hat{\lambda}_j \cdot (\hat{\lambda}_{j+1} - 1) = \hat{\lambda}_j \cdot \hat{\lambda}_{j+1} - 1$, we obtain an estimation for the reserve of origin year i

$$\hat{R}_i = \hat{x}_i\hat{y}_{n+2-i} + \ldots + \hat{x}_i\hat{y}_j = D_{i,n+1-i}(\hat{\lambda}_{n+2-i} \cdot \ldots \cdot \hat{\lambda}_n - 1).$$

This shows us that the solution of the quasi-likelihood equations of the over-dispersed Poisson model gives the same estimator for ultimate claim D_{in} as the chain-ladder algorithm (the distribution-free model of Mack).

Allowing for over-dispersion does not affect estimation of the parameters, but does have the effect of increasing their standard errors (England, Verrall, 2002). Over-dispersion is taken into account by estimating the unknown scale parameter ϕ as a part of the fitting procedure, using Pearson's residuals. Parameter ϕ is estimated by $\hat{\phi} = \frac{\chi^2}{df}$, where χ^2 is Pearson's statistic. Degrees of freedom $df = k - p$ consists the number of all past observations $k = \frac{n\cdot(n+1)}{2}$ in the run-off triangle and number of estimated parameters $p = 2n - 1$. Pearson's residuals can be calculated by

$$r_{iP} = \frac{C_i - \hat{C}_i}{\sqrt{\hat{C}_i}},$$

where $\hat{C}_i$ are fitted values obtained by backwards recursion using chain-ladder estimates. We use the result $\chi^2 = \sum_i (r_{iP})^2$ to obtain the estimation for ϕ:

$$\hat{\phi} = \frac{\sum_i \left(\frac{C_i - \hat{C}_i}{\sqrt{\hat{C}_i}}\right)^2}{df}. \tag{2.7}$$

Note that the index i used in this discussion denotes all the past observations in the run-off triangle.

The mean squared error of the prediction (MSEP) $\hat{C}_{ij}$ is given by

$$\begin{aligned} MSEP(\hat{C}_{ij}) &= E((C_{ij} - \hat{C}_{ij})^2) \\ &\approx (E(C_{ij}) - E(\hat{C}_{ij}))^2 + Var(C_{ij} - \hat{C}_{ij}) \\ &= Var(C_{ij}) + Var(\hat{C}_{ij}). \end{aligned}$$

The squared bias is small and can be left out as long as the estimators $\hat{C}_{ij}$, even if not unbiased, are asymptotically unbiased predictors of C_{ij}. The future observation C_{ij} and its forecast $\hat{C}_{ij}$ are independent random variables, thus the variance of their difference is just the sum of their variances.

We use the notation $\mu_{ij} = x_i y_j$ for simplification. Therefore, the equation (2.5), i.e. $\hat{C}_{ij} = \hat{\mu}_{ij} = \exp(\hat{\eta}_{ij})$ and the general form $Var(C_{ij}) = \phi x_i y_j$, are used in *delta method* for deriving $Var(\hat{C}_{ij})$. The idea of delta method is that for a function $g(\cdot)$ and a random variable X with mean $E(X) = \gamma$ and variance $Var(X) = \sigma^2$, the approximation $Var(g(X)) \approx (g'(\gamma))^2 \cdot \sigma^2$ holds. Using that method, we see

$$Var(\hat{C}_{ij}) \approx \left| \frac{\partial \hat{\mu}_{ij}}{\partial \hat{\eta}_{ij}} \right|^2 Var(\hat{\eta}_{ij}). \tag{2.8}$$

Since $\partial\hat{\mu}/\partial\hat{\eta} = \hat{\mu}$ in case of log-link and combining the equation (2.8) with the general form of variance, we get the MSEP of future payment C_{ij}:

$$MSEP(\hat{C}_{ij}) \approx \hat{\phi}\hat{\mu}_{ij} + \hat{\mu}_{ij}^2 Var(\hat{\eta}_{ij}),$$

where the first term denotes the process error and the last term denotes the estimaton error.

The estimation of MSEP for several cells (i, j) is more difficult since we obtain correlations from the estimates. Thus, the calculation of prediction errors for origin year reserve estimates and overall reserve estimates require more effort. The variance of the sum of predicted values is considered, taking account of

any covariances between predicted values. Making certain assumptions, the only covariances which need to be considered arise in the estimation variance. This should be easy to understand, since predicted values in each row are based on the same parameters and predicted values in the same column are based on the same parameters, so it leads to dependency.

If $\hat{C}_{i_1 j_1}$ and $\hat{C}_{i_2 j_2}$ are different estimated future payments, then

$$Cov(\hat{C}_{i_1 j_1}, \hat{C}_{i_2 j_2}) \approx \hat{\mu}_{i_1 j_1} \hat{\mu}_{i_2 j_2} \widehat{Cov}(\hat{\eta}_{i_1 j_1}, \hat{\eta}_{i_2 j_2}),$$

the same approximation holds in similar fashion for $\hat{C}_{ij_1}$ and $\hat{C}_{ij_2}$ (Kaas et al, 2008). We denote Δ as the triangle of predicted claims contributing to the reserve estimates. The reserve estimate for origin year i is given by sum of the predicted values in row i of Δ, thus

$$\hat{R}_i = \sum_{j \in \Delta_i} \hat{C}_{ij}.$$

The MSEP of the origin year i reserve is given by:

$$\begin{aligned}
MSEP(\hat{R}_i) &= E[(R_i - \hat{R}_i)^2] \\
&= \sum_{j \in \Delta_i} E[(C_{ij} - \hat{C}_{ij})^2] + \sum_{j_1, j_2 \in \Delta_i} E[(C_{ij_1} - \hat{C}_{ij_1})(C_{ij_2} - \hat{C}_{ij_2})] \\
&\approx \sum_{j \in \Delta_i} E[(C_{ij} - \hat{C}_{ij})^2] + \sum_{j_1, j_2 \in \Delta_i} Cov[C_{ij_1} - \hat{C}_{ij_1}, C_{ij_2} - \hat{C}_{ij_2}] \\
&= \sum_{j \in \Delta_i} E[(C_{ij} - \hat{C}_{ij})^2] + \sum_{j_1, j_2 \in \Delta_i} Cov(\hat{C}_{ij_1}, \hat{C}_{ij_2}) \\
&\approx \sum_{j \in \Delta_i} \hat{\phi} \hat{\mu}_{ij} + \sum_{j \in \Delta_i} \hat{\mu}_{ij}^2 Var(\hat{\eta}_{ij}) + 2 \sum_{\substack{j_1, j_2 \in \Delta_i \\ j_2 > j_1}} \hat{\mu}_{ij_1} \hat{\mu}_{ij_2} \widehat{Cov}(\hat{\eta}_{ij_1}, \hat{\eta}_{ij_2}).
\end{aligned}$$

The overall reserve estimate is defined as

$$\hat{R} = \sum_{i,j \in \Delta} \hat{C}_{ij} \tag{2.9}$$

and the MSEP of the overall reserve is given by:

$$\begin{aligned}
MSEP(\hat{R}) &= E[(R-\hat{R})^2] \\
&= \sum_{i,j\in\Delta} E[(C_{ij}-\hat{C}_{ij})^2] + \sum_{\substack{i_1,j_1\in\Delta \\ i_2,j_2\in\Delta}} E[(C_{i_1j_1}-\hat{C}_{i_1j_1})(C_{i_2j_2}-\hat{C}_{i_2j_2})] \\
&\approx \sum_{j\in\Delta_i} E[(C_{ij}-\hat{C}_{ij})^2] + \sum_{\substack{i_1,j_1\in\Delta \\ i_2,j_2\in\Delta}} Cov[C_{i_1j_1}-\hat{C}_{i_1j_1}, C_{i_2j_2}-\hat{C}_{i_2j_2}] \\
&= \sum_{j\in\Delta_i} E[(C_{ij}-\hat{C}_{ij})^2] + \sum_{\substack{i_1,j_1\in\Delta \\ i_2,j_2\in\Delta}} Cov(\hat{C}_{i_1j_1}, \hat{C}_{i_2j_2}) \qquad (2.10) \\
&\approx \sum_{i,j\in\Delta} \hat{\phi}\hat{\mu}_{ij} + \sum_{i,j\in\Delta} \hat{\mu}_{ij}^2 Var(\hat{\eta}_{ij}) + 2 \sum_{\substack{i_1,j_1\in\Delta \\ i_2,j_2\in\Delta \\ (i_1,j_1)\neq(i_2,j_2)}} \hat{\mu}_{i_1j_1}\hat{\mu}_{i_2j_2}\widehat{Cov}(\hat{\eta}_{i_1j_1}, \hat{\eta}_{i_2j_2}).
\end{aligned}$$

The first term in the origin year prediciton error and in overall prediction error is simply the appropriate sum of the process variances and the remaining terms relate to the estimation variance.

2.3 Negative Binomial Model

In the context of GLMs the first stochastic version of the chain-ladder method that can be applied in the presence of negative incremental claim values is defined as a generalized linear model with an over-dispersed Poisson distribution (Renshaw and Verrall 1998). Overdispersion is achieved if the scale parameter $\phi > 1$. The model reproduces the estimates of the basic chain-ladder method and the estimates of the parameters $\hat{c}$, $\hat{\alpha}_i$, $\hat{\beta}_j$ are obtained by using a quasi-likelihood approach.

Another model, which can be considered as chain-ladder "type", is negative binomial model. This model is derived from the Poisson model and, as such, is very closely related, but with a different parameterisation.

Remark 2.4. *The negative binomial model was first derived by Verrall (2000), by integrating out the row parameters from the Poisson model.*

The ODP model has origin year parameters and development year parameters, but the parameters in the negative binomial model relate to development years only. The parameters of the negative binomial model appear to be more 'like' the chain-ladder development factors, so it makes it more intuitive for the chain-ladder method. Either incremental or cumulative data can be considered in this model. The incremental claim C_{ij} has an over-dispersed negative binomial distribution, with mean

$$E(C_{ij}) = (\lambda_j - 1)D_{i,j-1},$$

and variance

$$Var(C_{ij}) = \phi\lambda_j(\lambda_j - 1)D_{i,j-1},$$

where the parameters $(\lambda_j : j = 1, \ldots, n-1)$ are the typical chain-ladder development factors defined in the section 1 and the D_{ij} (where observed) are considered known. Here ϕ is an overdispersion parameter as in the Poisson model and again, makes the distribution 'over-dispersed'. This label is dropped for convenience when referring to the negative binomial model.

Note the recursive expression $D_{ij} = D_{i,j-1} + C_{ij}$, where $D_{i,j-1}$ is assumed to be known. This leads to a model for cumulative data. More precisely, D_{ij} has an over-dispersed negative binomial distribution with mean

$$E(D_{ij}) = \lambda_j D_{i,j-1}, \tag{2.11}$$

and variance

$$Var(D_{ij}) = \phi\lambda_j(\lambda_j - 1)D_{i,j-1},$$

noting that the variance holds only where $D_{i,j-1}$ is known. The chain-ladder is easy to apply due to parameterisation as in (2.11), with the mean in a particularly simple form. It happens because the column parameters represent separate factors, and can be replaced by another set of factors, the development factors.

We proceed considering the model for incremental data, with

$$E(C_{ij}) = (\lambda_j - 1)D_{i,j-1} \text{ and } Var(C_{ij}) = \phi\lambda_j(\lambda_j - 1)D_{i,j-1}, \tag{2.12}$$

where the D_{ij} (where observed) are assumed to be known. Next, we specify a GLM with log-link and negative binomial error structure. Using already familiar notation, we write

$$E(C_{ij}) = \mu_{ij} = (\lambda_j - 1)D_{i,j-1},$$

where taking logarithms gives

$$\log(\mu_{ij}) = \log(\lambda_j - 1) + \log(D_{i,j-1}).$$

The first term $\log(\lambda_j - 1)$ in the sum can be given as

$$\log(\lambda_j - 1) = c + \alpha_{j-1} \text{ with } \alpha_1 = 0,\ j \geq 2, \tag{2.13}$$

which defines the negative binomial model in terms of GLM as follows:

$$\log(\mu_{ij}) = c + \alpha_{j-1} + \log(D_{i,j-1}). \tag{2.14}$$

The $\log(D_{i,j-1})$ terms are derived from the known values $D_{i,j-1}$, and are specified as offsets in the model. One can notice from (2.14) that there is the parameter only for development years, and in comparison with ODP, the row parameters no longer appear.

This model is now formulated as a recursive model. We require recursive procedures to obtain the estimation variance and process variance. The following results and derivations can be found in England & Verrall (2002). For the origin year reserve estimate, the ultimate claims D_{in} are the cumulative claims in the last development years. The reserve estimate in origin year i is $R_i = D_{in} - D_{i,n+1-i}$,

where $D_{i,n+1-i}$ is the last observed (known) value in the run-off triangle for row i. Therefore,

$$Var(R_i) = Var(D_{in}) \text{ and } Var(\hat{R}_i) = Var(\hat{D}_{in}).$$

The origin year process and estimation variance can be estimated by considering $Var(D_{in})$ and $Var(\hat{D}_{in})$ respectively. First of all, we need an estimation for λ_j factors. We use the general rule of the approximate variance of a function, that is

$$Var[f(X)] \approx (f'(E[X]))^2 Var(X).$$

Thus, the estimates of the development factors can be obtained from the parameter estimates using equation (2.13), and their approximate standard errors can be obtained with:

$$Var(\hat{\lambda}_j) = Var(\hat{\lambda}_j - 1) \approx (\exp(\hat{c} + \hat{\alpha}_{j-1}))^2 Var(\hat{c} + \hat{\alpha}_{j-1}),\ j \geq 2.$$

Next, we consider the process variance $Var(D_{in})$. Since the model is formulated as recursive models, the calculation of the process variance involves estimating the variance of a k-steps-ahead forecast (where $k = i - 1$), using standard results from the analysis of conditional distributions. Formally, the variance of a k-steps-ahead forecast is given by:

$$Var(D_{i,j+k-1} \mid C_1, C_2, \ldots, C_{j-1}) = \phi\lambda_j\lambda_{j+1}\ldots\lambda_{j+k-1}(\lambda_j\lambda_{j+1}\ldots\lambda_{j+k-1} - 1)D_{j-1}.$$

In our context with $n = j + k - 1$ the process variance is calculated from:

$$Var(D_{in}) \approx \phi D_{i,n+1-i} \prod_{k=n+2-i}^{n} \hat{\lambda}_k \left(\prod_{k=n+2-i}^{n} \hat{\lambda}_k - 1 \right). \qquad (2.15)$$

The estimation error can also be calculated recursively. We require $Var(\hat{D}_{in}|D_{i,n+1-i})$, which is the variance of the predicted row total, given the latest cumulative claims.

Note that this is the same as the variance of the sum of incremental predicted values, which provides an alternative way of calculating the estimation error, but gives the same result:

$$\begin{aligned} Var(\hat{D}_{in}|D_{i,n+1-i}) &= Var(\hat{\lambda}_{n+2-i}\ldots\hat{\lambda}_n D_{i,n+1-i} \mid D_{i,n+1-i}) \\ &= D^2_{i,n+1-i} Var(\hat{\lambda}_{n+2-i}\ldots\hat{\lambda}_n \mid D_{i,n+1-i}). \end{aligned}$$

To simplify the notation, the explicit conditioning notation is dropped. Thus, the estimation variance is computed from:

$$Var(\hat{D}_{in}) \approx Var\left(D_{i,n+1-i}\prod_{k=n+2-i}^{n}\hat{\lambda}_k\right) = D^2_{i,n+1-i}Var\left(\prod_{k=n+2-i}^{n}\hat{\lambda}_k\right)$$

The overall reserve estimation and process variance can be estimated by considering $Var(R)$ and $Var(\hat{R})$ respectively, where $R = \sum_{i=2}^{n} R_i$. The overall reserve process variance is the sum of the process variances of individual origin year reserves, assuming independence between years. For the overall reserve, the estimation variance is given by:

$$Var(\hat{R}) \approx \sum_{i=2}^{n} Var(\hat{D}_{in}) + 2\sum_{\substack{i=2 \\ j>i}}^{n} Cov(\hat{D}_{in}, \hat{D}_{jn}),$$

i.e. the estimation variance of overall reserves is the sum of the estimation variances of individual origin year reserves, with an additional component to take account of the covariance between years induced by dependence on the same parameters.

The reserve estimates can be calculated from the development factors, and the prediction error of the reserves can be calculated using given equations above for process and estimation variances.

Since the negative binomial model is derived from the Poisson model, the predictive distributions are basically the same and give identical predicted values.

Estimation variance is larger for ODP than negative binomial, but process variance is larger for negative binomial than ODP. The practical implication of this is that it does not matter which model is fitted, the results will be the same. Essentially, ODP and negative binomial model only differ in the way the models are parameterised.

With enough negative incremental claims, it is possible that some of the development factors λ_j (one would be enough) become less than one. In that case the variance would not exist and the model cannot be applied. Therefore, with a sufficient number of negative incremental claims, it is recommended to use a Normal approximation, and the chain-ladder results can still be reproduced.

2.4 Normal approximation to negative binomial model

The negative binomial model breaks down if the development factor $\lambda_j < 1$, since this produces negative variance. $\lambda_j < 1$ means that the sum of incremental claims in column j is negative. Among the models considered there are several that can handle negative values, another one is Normal approximation to negative binomial model. The Normal model has the advantage that it can provide estimates for a wide range of data sets, and is less affected by the presence of negatives.

Firstly, it is possible to replace the negative binomial by a Normal distribution with the same mean as it was in negative binomial model, but the variance is altered to handle the case of $\lambda_j < 1$. Preserving as much of $Var(C_{ij}) = \lambda_j(\lambda_j - 1)D_{i,j-1}$ as possible, the variance is expected to be proportional to $D_{i,j-1}$, with the constant of proportionality depending on j. Thus, using a Normal approximation for the distribution of incremental claims, C_{ij} is approximately Normally distributed with mean

$$E(C_{ij}) = D_{i,j-1}(\lambda_j - 1)$$

and variance

$$Var(C_{ij}) = \phi_j D_{i,j-1},$$

that the D_{ij} (where observed) are considered known. The normal approximation to the negative binomial can also be fitted using cumulative data. So, if cumulative claims are forecast, then distribution of D_{ij} is assumed to approximately follow a normal distribution with mean

$$E(D_{ij}) = \lambda_j D_{i,j-1} \tag{2.16}$$

and variance

$$Var(D_{ij}) = \phi_j D_{i,j-1}, \tag{2.17}$$

where D_{ij} is also assumed to be known and a recursive approach ($D_{ij} = D_{i,j-1} + C_{ij}$) is required for estimating the variance.

We consider using cumulative data. Denoting $w_{ij} = D_{i,j-1}$ and dividing (2.16) and (2.17) by w_{ij} gives

$$E\left(\frac{D_{ij}}{w_{ij}}\right) = \lambda_j \text{ and } Var\left(\frac{D_{ij}}{w_{ij}}\right) = \frac{\phi_j}{w_{ij}}.$$

Analysing this model, the focus will be on the quantities $f_{ij} = \frac{D_{ij}}{w_{ij}}$, the individual development factors, which are approximately independently and normally distributed within the development year j. Also, the variance components ϕ_j depend on development year j. The variable w_{ij} has been introduced since a weighted linear model will be used in estimating the unknown individual development factors. The variance components ϕ_j depend on the development year j, but usually in a GLM the dispersion parameter ϕ is constant for all observations. The weights in this discussion are given with $W_{ij} = \frac{w_{ij}}{\phi_j}$. The weights are inversely proportional with the variance, so that data with a greater variance is less weighted.

The linear model can be given as follows

$$E(f_{ij}) = c + \alpha_{j-1} \text{ with a restriction } \alpha_1 = 0,\ j \geq 2.$$

The linear predictor above depends only on development period j, and the fitted values $\hat{f}_{ij}$ are estimates of the development factors λ_j. It is assumed f_{ij} to be independent and since also normally distributed, the link function is only the identity function. Both the parameters ϕ_j and f_{ij} need to be estimated, they are modelled as part of the joint modelling, for which the technique is described in Renshaw(1994a) and England & Verrall(2002). We follow the description given by England & Verrall(2002).

The joint modelling process proceeds by providing initial random positive values for parameters ϕ_j, and fitting a weighted GLM to the f_{ij} with weights $W_{ij} = \frac{w_{ij}}{\phi_j}$. Having obtained the fitted values, the square of the residuals r_{ij} are calculated by

$$r_{ij}^2 = w_{ij}(f_{ij} - \hat{f}_{ij})^2.$$

These squared residuals are used as the response in a second model, in which the predictor depends on development year j only. The fitted values $\hat{r}_{ij}^2$ from this second model are used to update the values of ϕ_j. Then have to refit the first model after updating weights W_{ij} to reflect the revised estimates of ϕ_j. This completes the joint modelling process, from which the estimates of λ_j and ϕ_j can be obtained.

Both last models considered, the negative binomial and the Normal approximation, are recursive models. The theory underlying the calculation of prediction errors for this, Normal approximation, model is the same as for the negative binomial model in previous subsection. Mack's model takes similar recursive approach and is closely related to the Normal approximation to the negative binomial. One can note the similarity in comparison of variances of both models,

where the unknown scale parameters ϕ_j of the Normal approximation can be replaced by σ_j^2 in Mack's model.

2.5 Log-normal model

We continue this section with a non-recursive model, with log-normal model. In log-normal model, the first step is to transform the incremental claims by taking their (natural) logarithm and using a normal distribution.

The log-normal class of models are given as

$$\log(C_{ij}) \sim N(\mu_{ij}, \sigma^2),$$

i.e.

$$E(\log(C_{ij})) = \mu_{ij} \text{ and } Var(\log(C_{ij})) = \sigma^2.$$

The normal responses $\log(C_{ij})$ are assumed to decompose (additively) into a deterministic non-random component with mean $\mu_{ij} = \eta_{ij}$ and Normally distributed random error components about a zero mean.

Using the chain-ladder type predictor structure, we have:

$$\eta_{ij} = c + \alpha_i + \beta_j$$

The use of the logarithmic transform immediately imposes a limitation on this class of models that incremental amounts must be positive. This model usually produces predicted values close to predictions from the chain-ladder method, but it can not be guaranteed, and there may be material differences.

Having obtained estimated for the parameters in the linear predictor η_{ij}, and the process variance σ^2, the fitted values on a log scale are obtained by forming the appropriate sum of estimates. To obtain the estimates for the mean on the untransformed scale is not that simple. We cannot just exponentiate the linear

predictor, since that would give an estimate of the median. Therefore, the fitted values on the untransformed scale are given by:

$$\hat{C}_{ij} = \exp(\hat{\eta}_{ij} + \frac{1}{2}\hat{\sigma}^2_{ij}), \tag{2.18}$$

which is in the standard form of the expected value of a log-normal distribution and where

$$\hat{\sigma}^2_{ij} = Var(\hat{\eta}_{ij}) + \hat{\sigma}^2.$$

In fact, the $\hat{\sigma}^2_{ij}$ terms are the prediction variance of the linear predictor and are calculated as the sum of the variance of η_{ij} and the underlying process variance, so the variance component includes the estimation and process error.

Remark 2.5 (England, Verrall, 2002). *It is important to note that the reason why a variance component in equation (2.18) is needed for the log-normal model, but not for the over-dispersed Poisson, is because the incremental claims, themselves, are used as the response with the ODP model, but with the log-normal model, the logarithm of the incremental claims is used as the responses.*

The prediction variance of future incremental claims C_{ij} is given by:

$$MSEP(\hat{C}_{ij}) = \hat{C}^2_{ij}(\exp(\hat{\sigma}^2_{ij}) - 1)$$

which is in the standard form of the variance of a log-normal distribution. The prediction error of origin year and overall reserve estimates requires more effort, like the ODP model. The variance of the sum of predicted values is considered, taking account any covariances between predicted values. With already familiar notation (from the ODP subsection), we denote the triangle of predicted claims contributing to the reserve estimates by Δ. The following results can be found in England & Verrall (2002). The reserve estimate in origin year i is given by summing the predicted values in row i of Δ, i.e.

$$\hat{R}_i = \sum_{j \in \Delta_i} \hat{C}_{ij}.$$

The mean squared error of prediction of the origin year reserve is given by:

$$MSEP(\hat{R}_i) \approx \sum_{j \in \Delta_i} MSEP(\hat{C}_{ij}) + 2 \sum_{\substack{j_1, j_2 \in \Delta \\ j_2 > j_1}} \hat{C}_{ij_1} \hat{C}_{ij_2} (\exp(Cov[\hat{\eta}_{ij_1}, \hat{\eta}_{ij_2}]) - 1).$$

The total reserve estimate, summing the predicted values in row i and in column j of Δ, is given by:

$$\hat{R} = \sum_{i,j \in \Delta_i} \hat{C}_{ij}. \tag{2.19}$$

The MSEP of the total reserve is given by:

$$MSEP(\hat{R}) \approx \sum_{i,j \in \Delta_i} MSEP(\hat{C}_{ij}) + 2 \sum_{\substack{i_1, j_1 \in \Delta \\ i_2, j_2 \in \Delta \\ i_1 j_1 \neq i_2 j_2}} \hat{C}_{i_1 j_1} \hat{C}_{i_2 j_2} (\exp(Cov[\hat{\eta}_{i_1 j_1}, \hat{\eta}_{i_2 j_2}]) - 1). \tag{2.20}$$

In log-normal model, the same range of predictor structures is available as before. It is important to note the component of variance in the mean on the untransformed scale. For example, the estimate of the expected reserves given by the log-normal model can be far from the chain-ladder estimate due to the inclusion of the variance component in equation (2.18), which can have a significant effect when the underlying variability is large. But when focus is on the mean, ignoring the variance component is invalid.

2.6 Gamma model

Mack (1991) proposed further model with a multiplicative parametric structure for the mean incremental claims amounts which are modelled as Gamma response variables. As Renshaw & Verrall (1998) note, the same model can be fitted using the GLM described in over-dispersed Poisson model, but in which the incremental claim amounts are modelled as independent Gamma response variables, with a logarithmic link function and the same linear predictor, and just replacing $Var(C_{ij}) = \phi \mu_{ij}$ by $Var(C_{ij}) = \phi \mu_{ij}^2$. As it was with log-normal model,

the predicted values provided by Gamma model are usually close to chain-ladder estimates, but it cannot be quaranteed.

Remark 2.6 (**England, Verrall, 1998).** *The Gamma model implemented as a generalised linear model gives exactly the same reserve estimates as the Gamma model implemented by Mack (1991), which is comforting rather than surprising.*

To obtain predictions and prediction errors for the Gamma model simply requires a small change in the ODP model. The Gamma model is given with the mean

$$E(C_{ij}) = \mu_{ij},$$

and with variance

$$Var(C_{ij}) = \phi(E(C_{ij}))^2 = \phi\mu_{ij}^2,$$

so the variance in this model is proportional to the mean squared, not proportional to the mean as in the case of ODP model.

Remark 2.7. *We need to impose that each incremental value should be non-negative if we work with gamma (Poisson) models. This restriction can be overcome using a quasi-likelihood approach.*

Using the chain-ladder type predictor structure

$$\eta_{ij} = c + \alpha_i + \beta_j,\ \alpha_1 = \beta_1 = 0,$$

$$\log(\mu_{ij}) = \eta_{ij},$$

it is straightforward to obtain parameter estimates and predicted values using GLM. Like stated before, the Gamma model relates to ODP model, therefore estimating the prediction errors of future payments and reserves is comparatively simple by making just the appropriate change to the corresponding process error

components in ODP model. The mean squared error of the prediction of $\hat{C}_{ij}$ given by

$$MSEP(\hat{C}_{ij}) \approx Var(C_{ij}) + Var(\hat{C}_{ij})$$

in ODP model is also valid in Gamma model. Thus, the prediction variance of future incremental claims C_{ij} is:

$$MSEP(\hat{C}_{ij}) \approx \hat{\phi}\hat{\mu}_{ij}^2 + \hat{\mu}_{ij}^2 Var(\hat{\eta}_{ij}),$$

the MSEP of the origin year i reserve is as follows

$$MSEP(\hat{R}_i) \approx \sum_{j\in\Delta_i} \hat{\phi}\hat{\mu}_{ij}^2 + \sum_{j\in\Delta_i} \hat{\mu}_{ij}^2 Var(\hat{\eta}_{ij}) + 2 \sum_{\substack{j_1,j_2\in\Delta_i \\ j_2>j_1}} \hat{\mu}_{ij_1}\hat{\mu}_{ij_2}\widehat{Cov}(\hat{\eta}_{ij_1}, \hat{\eta}_{ij_2}).$$

and the MSEP of the overall reserve is

$$MSEP(\hat{R}) \approx \sum_{i,j\in\Delta} \hat{\phi}\hat{\mu}_{ij}^2 + \sum_{i,j\in\Delta} \hat{\mu}_{ij}^2 Var(\hat{\eta}_{ij}) + 2 \sum_{\substack{i_1,j_1\in\Delta \\ i_2,j_2\in\Delta \\ (i_1,j_1)\neq(i_2,j_2)}} \hat{\mu}_{i_1j_1}\hat{\mu}_{i_2j_2}\widehat{Cov}(\hat{\eta}_{i_1j_1}, \hat{\eta}_{i_2j_2}). \tag{2.21}$$

Recall that the first term in the accident year and overall prediction errors is simply the appropriate sum of the process variances. The remaining terms relate to the estimation variance.

3 Comparison of the methods

This section gives an overview of main differences (or similarities) that can be noticed among the methods introduced in the thesis. Lot of attention has been paid to chain-ladder method and to describe the stochastic model underlying it. As stated before, only ODP and distribution-free model of Mack could be considered as the chain-ladder "type" model reproducing reserve estimates given by the chain-ladder technique, therefore we examine closely these two models and bring out the main differences. Also, since among the introduced models only few can cope with negative incremental claims, a short discussion about negative values in the run-off triangle is given.

3.1 Mack's stochastic model and the over-dispersed Poisson model

We have given the outline of several reserving methods in the thesis. We considered the chain-ladder method, which is a deterministic algorithm implying nothing about the variability of the actual outcome. Thus the stochastic models are needed in order to find the variability of the estimate. Therefore we introduced distribution-free model of Mack (DFCL), which is called *the* stochastic model underlying the chain-ladder algorithm. Also, the (over-dispersed) Poisson model is considered as the chain-ladder "type" model reproducing reserve estimates given by the chain-ladder technique. Nevertheless, the ODP and distribution-free model of Mack are different and they yield the same predictions only in a special but common situations. Both methods yield the same estimators $\hat{R}_i$ and $\hat{D}_{in}$ as the deterministic chain-ladder algorithm, thus one can think whether one model is derived from other. We would like to point out the main differences of these two models, using the arguments formulated by Mack & Venter (1999), each of which

shows that the models are different:

1) ODP has more parameters than DFCL. In ODP, one of the parameters is redundant because replacing x_i with x_i/c and y_j with $y_j \cdot c$ yields the same model. Therefore ODP has $2n-1$ parameters whereas DFCL has only $n-1$ parameters $\lambda_1, \ldots, \lambda_{n-1}$.

2) ODP and DFCL have different independence assumptions. Either model assumes that the accident years (rows) are independent. In addition, the assumption (ODP3) requires all increments C_{ij} within each accident year to be independent. DFCL implies that all C_{ij} within the same accident year are correlated, i.e. given $D_{i,j-1}$, we have

$$C_{i,j+1} = D_{i,j+1} - D_{ij} = D_{ij}(\lambda_{j+1} - 1) = (D_{i,j-1} + C_{ij})(\lambda_{j+1} - 1),$$

where can be seen that $C_{i,j+1}$ and C_{ij} are correlated.

3) The fitted values $\hat{C}_{ij}$ or $\hat{D}_{ij}$ and therefore also the residuals $r_{ij} = \hat{C}_{ij} - C_{ij}$ for ODP are different from residuals of DFCL. The fitted value $\hat{D}_{ij}$ for $j \leq n+1-i$ for the DFCL model is

$$\hat{D}_{ij} = D_{i,j-1}\hat{\lambda}_j \text{ or } \hat{C}_{ij} = D_{i,j-1}(\lambda_j - 1).$$

There are no fitted values for the first column $D_{i1} = C_{i1}$, $i = 1, \ldots, n$. The fitted value $\hat{D}_{ij}$, $j \leq n+1-i$, for the ODP model is

$$\hat{C}_{ij} = \hat{x}_i\hat{y}_j = D_{i,n+1-i}((\hat{\lambda}_{j+1} \cdot \ldots \cdot \hat{\lambda}_{n+1-i})^{-1} - (\hat{\lambda}_j \cdot \ldots \cdot \hat{\lambda}_{n+1-i})^{-1}),$$

which yields

$$\hat{D}_{ij} = D_{i,n+1-i}/(\hat{\lambda}_{j+1} \cdot \ldots \cdot \hat{\lambda}_{n+1-i}),$$

and differs from DFCL, because e.g. for $j = n+1-i$ we obtain $D_{i,n-i} \cdot \hat{\lambda}_{n+1-i}$ as fitted value for DFCL and $D_{i,n+1-i}$ as fitted value for ODP which are different for $i > 1$ because $\hat{\lambda}_{n+1-i}$ depends not only on $D_{i,n+1-i}/D_{i,n-i}$ of accident year i

but also on the corresponding ratios of all older accident years.

4) The simulated future emergence is different. To simulate the future values of C_{ij} for ODP, one should calculate the expected values, and then add randomness. For DFCL, the first new diagonal could be simulated this way, but for the second new diagonal the simulated cumulative value for the first diagonal would have to be multiplied by the development factor to get the mean value for the second diagonal. This mean thus includes the random component simulated for the first diagonal, which is not the case for the ODP simulation.

3.2 Negative incremental values

Incremental loss data triangles frequently contain negative values. Typically these negative values will be the result of salvage recoveries, payments from third parties, total or partial cancellation of outstanding claims due to initial overestimation of the loss or to a possible favorable jury decision in favor of the insurer, rejection by the insurer, or just plain errors (de Alba, 2006). There have been proposed different solutions on how to handle negative incremental claims. Mack (1994) suggested to assign unwanted incremental claims a missing value status and to subsequently leave them out of the analysis.

Many stochastic claims reserving models have dealt with negative values in incremental loss data triangles by using the *constant-solution*. This solution adds a constant to all incremental claims prior to analysis, which forces them to be positive definite (greater than zero). Afterwards the same constant is removed from the incremental forecasts (Verrall, Li, 1993). This method provides suitable results as long as there are not too many negative claims. On the other hand, this procedure makes the variability of the result depend on the constant added earlier, which cannot be considered reasonable (Kunkler, 2006).

Another possibility is to use the model which can handle negative claims. The

chain-ladder technique and Mack's model may produce reserve estimates even when there are negative values as long as the cumulative claims are positive. If the distribution is specified in the model, it needs to be defined for negative as well as for positive numbers. However, many methods can break down in the presence of sufficient number of negative incremental claims. Clearly, one suitable candidate is the normal distribution, which is defined for both positive and negative numbers.

Neither the Poisson model nor the negative binomial model can contain negative incremental claims if the usual maximum likelihood estimator of the parameters is used. As long as the sum of the incremental claims belonging to one development year is not negative, this problem can be solved by using a quasi log-likelihood (Renshaw, Verrall, 1998). As an alternative to the negative binomial model, a normal approximation to the negative binomial model was used since this would solve the problem with negative incremental claims. Nevertheless, it is not recommended to use normal approximation since more parameters need to be estimated (the variance factors). The log-normal model assumes definite positive incremental claims due to a logarithmic transformation performed on the incremental loss data. Gamma model relates to ODP model, therefore in case of Gamma model incremental values should be non-negative as well.

3.3 Conclusions

The relationships between some of the models were explored. The DFCL model agrees with the chain-ladder algorithm in every identified area, whereas the ODP does not. For instance, the chain-ladder algorithm does not compute any parameters other than the development factors, and if one would want to measure goodness-of-fit for the chain-ladder algorithm, one would calculate the fitted values as the development factors times the previous cumulative losses. The ODP differs from the chain-ladder algorithm in these aspects.

The normal approximation to the negative binomial model and Mack's model are two possible models in case of negative values in the data. These two models provide nearly identical results, and the normal approximation can be seen as an underlying model of Mack's model. However, since the normal approximation model is based on a GLM, the normal approximation to the negative binomial model offers greater flexibility in applied calculations than Mack's model.

The negative binomial model is derived from the over-dispersed Poisson model. Hence, the restriction of negative incremental claims for the overdispersed Poisson model are the same as for the binomial model. The negative binomial model has fewer parameters to be estimated than the Poisson model, but predicted values for both models are identical and their predictive distributions are essentially the same (England, Verrall, 2002).

Reserving is a practical data analysis exercise and if one model is suitable for several data, it does not mean that the model can be applied in all situations. Therefore it is suggested to learn from the data and experiment with different approaches. Clearly, different stochastic methods will give different results. The decision about reserve estimate should be a compromise between actuary's experience and outcome of statistical model.

4 The practical implementation

The following practical implementation is based on the real data from the insurance company. The data considered describes the paid out claims. For the implementation we use Mack's model, over-dispersed Poisson model, log-normal model and Gamma model. The data set used is as follows.

	1	2	3	4	5	6	7	8	9	10
2000	4 734 994	1 885 305	281 240	504 341	524 449	365 049	100 761	32 449	3 697	56 901
2001	4 344 093	1 783 774	243 849	339 985	49 482	178 961	508 272	78 125	1 022	
2002	5 288 867	1 795 855	303 246	351 320	316 038	33 501	88 774	31 102		
2003	5 357 617	2 548 383	336 749	403 501	348 378	236 017	12 982			
2004	5 737 732	2 574 724	971 320	280 140	226 212	152 127				
2005	5 635 064	2 758 392	241 734	268 113	429 503					
2006	6 629 504	3 045 252	356 119	200 420						
2007	6 824 829	2 669 579	166 400							
2008	8 116 439	3 428 535								
2009	10 660 074									

Table 1: *Full run-off triangle for paid out claims*

We analyse the full triangle. The results of estimating reserves are given with the standard error in the following table.

	$\hat{R}$	$se(\hat{R})$	Prediction error %
Mack	13 405 108	1 852 202	13%
ODP	13 405 108	1 985 629	15%
Log-normal	17 430 668	12 851 133	74%
Gamma	12 142 220	5 411 186	44%

Table 2: *Reserve estimations and standard errors*

The overall reserve estimate of Mack model is obtained recursively by chain-ladder technique and for overall prediction error (1.4) is used. Notice that the overall reserve estimates of Mack model and ODP are identical and there is only small difference in prediction errors. For ODP model estimations, formulas (2.9) and (2.10) are used. Also cannot note high uncertainty in reserve estimates, the prediction errors of both models are quite reasonable. Therefore, theoretical statement that reserve estimates of Mack and ODP are identical is indeed valid in practical implementation.

For log-normal estimations, the equations (2.19) and (2.20) were used. It is clear from the Table 2 that the estimate of the expected reserves given by the log-normal model is far from the chain-ladder estimate, as is the overall prediction error, which, even as a percentage of the reserve estimate, is approximately 6 times larger. This can be explained by the inclusion of the variance component in equation (2.18), which can have a significant effect when the underlying variability is large.

Fitting the Gamma model with a chain-ladder type predictor, gives similar, but not identical, reserve estimate as results obtained by Mack model and ODP. However, the scale parameter in Gamma model is $\phi = 0.3217705$ and in ODP model it was $\phi = 95\,229$. The scale parameter for Poisson was estimated from the equation (2.7). Since in Gamma model the variance is proportional to the mean squared, the formula for ϕ is $\hat{\phi} = \frac{\sum_i \left(\frac{C_i - \hat{C}_i}{\hat{C}_i}\right)^2}{df}$. The overall prediction error is far from the ODP and Mack's model results. This kind of larger variability can be explained only by the underlying data. As one can see, the data considered is inconvenient (see Table 1), i.e. the large fluctuation of the values in triangle is obvious.

It is clear, that theoretical results given in previous sections hold also in practical situation. The Gamma model gave the smallest reserve estimate, but the vari-

ability of the model is higher than in ODP and Mack model. We also present the ratios of process and estimation variance, which is in the Poisson case 0.4787986, in log-normal model it is 0.4913061 and in Gamma model the result is 0.3430256, so the model outcomes can be considered reliable. Hence, it is up to an actuary which result should be taken into account when making decisions in setting up the fund for reserves.

Stohhastilised reservide arvutamise meetodid kahjukindlustuses

Magistritöö

Liivika Tee

Kokkuvõte

Kindlustusfirma juhindub eesmärgist tagada ettevõtte kasumlikkus ning solventsus ehk maksejõulisus. Solventsuse tagamiseks igal ajahetkel tulevikus peab kindlustusselts oskama hinnata oma tuleviku väljaminekuid. Kindlustuslepingunõuete täitmiseks on tarvis määrata tehnilised eraldised ehk reservid. Reservide hindamisega tegeleb aktuaar, kelle ülesandeks on statistilisi meetodeid ning erialast kogemust kasutades leida parim hinnang reservidele.

Käesoleva magistritöö eesmärgiks oli anda ülevaade erinevatest reservihindamise meetoditest. Põhjalikult kirjeldati laialt levinud ahel-redel meetodit, mis on oma lihtsuse tõttu üks enimkasutatavaid reservide hindamise meetodeid. Kuna ahel-redel meetodiga on võimalik saada ainult punkthinnang reservidele, siis ülejäänud osa tööst keskendus erinevate stohhastiliste meetodite uurimisele.

Esimeses peatükis räägiti lähemalt ahel-redel meetodist, selle omadustest ning reservide arvutamistehnikast. Kuna tihti ollakse huvitatud rohkemast kui ainult punkthinnang, siis peatükk jätkus Macki jaotusvaba mudeli uurimisega, mis võimaldab meil ka hinnangu varieeruvust hinnata. Macki jaotusvaba mudelit nimetataksegi ahel-redel meetodile aluseks olev stohhastiline mudel. Ahel-redel meetodit stohhastilise mudelina kujutades on võimalik leida reservidele vahemikhinnanguid. Kuna mudel on jaotusvaba, siis ei saa kasutada jaotuse karakteristikuid, seega kasutatakse varieeruvuse kirjeldamiseks ruutkeskmist viga.

Teine peatükk kätkeb endas erinevate parameetrilisite mudelite kirjeldust üldistatud lineaarsete mudelite terminoloogia alusel. Esitati multiplikatiivne mudel, millest lähtuvalt mudelid konstrueeriti. Toodi välja valemid reservihinnagute ning varieeruvuse arvutamiseks kõigi esitatud mudelite puhul.

Tähelepanu koondus suurelt Macki jaotusvaba ja ülehajuvusega Poissoni mudelile. Kuna mõlemad mudelid annavad identseid tulemusi ahel-redel meetodi tulemustega, siis kolmandas peatükis on antud ka nimetatud mudelite põhjalikum võrdlus. Lisaks arutleti, kuidas saada hakkama meetodite valikut kitsendava asjaoluga - kui arengukolmnurgas olevate väärtuste seas leidub negatiivseid väärtusi.

Teoreetiliste tulemuste kontrollimiseks toodi viimases peatükis näide reaalsete andmetega. Selleks katsetati mõningaid eelpool vaadeldud mudeleid ning üritati selgitada mudelite tulemusi. Vaadeldi mudelite poolt antud reservihinnanguid ning standardviga.

References

[1] De Alba, E. (2006). Claims Reserving When There Are Negative Values In The Runoff Triangle: Bayesian Analysis Using The Three-Parameter Log-Normal Distribution. *North American Actuarial Journal*, **3** (10), 45-59.

[2] England, P.D., Verrall, R.J. (1998). Standard errors of prediction in claims reserving: a comparison of methods. Analytic and bootstrap estimates of prediction errors in claims reserving. [WWW] http://www.actuaries.org.uk/system/files/documents/pdf/0459-0478.pdf

[3] England, P.D., Verrall, R.J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, **8** (3), 443-518.

[4] Holton, Glyn A. (2013). *Value-at-Risk: Theory and Practice*, 2nd ed. e-book at http://value-at-risk.net.

[5] Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R*. Springer, Berlin, Heidelberg.

[6] Kunkler, M. (2006). Modelling negatives in stochastic reserving models. *Insurance: Mathematics and Economics*, **38**, 540-555.

[7] Käärik, M. (2006). *Kahjukindlustusmatemaatika*. Loengukonspekt.

[8] Mack, T. (1991). A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves. *Astin Bulletin*, **21**, 93-109.

[9] Mack, T. (1993). Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *Astin Bulletin*, **23** (2), 213-225.

[10] Mack, T. (1994). Which Stochastic Model is Underlying the Chain Ladder Method. *Insurance: Mathematics and Economics*, **15** (2-3), 133-138.

[11] Mack, T., Venter, G. (1999). A Comparison of Stochastic Models that Reproduce Chain Ladder Reserve Estimates. [WWW] http://www.actuaries.org/ASTIN/Colloquia/Tokyo/Mack_Venter.pdf

[12] Renshaw, A.E., Verrall, R.J. (1998). A stochastic model underlying the chain ladder technique. *British Actuarial Journal*, **4** (4), 903-923.

[13] Schmidt, K. (2002). A note on the overdispersed Poisson family. *Insurance: Mathematics and Economics*, **30**, 21-25.

[14] Verrall, R.J. (1994). Statistical Methods for the Chain Ladder Technique. [WWW] http://www.casact.org/pubs/forum/94spforum/94spf393.pdf

[15] Verrall, R.J. (2000). An investigation into stochastic claims reserving models and the chainladder technique. *Insurance: Mathematics and Economics*, **26**, 91-99.

[16] Verrall, R.J., Li, Z. (1993). Negative incremental claims: Chain ladder and linear models. *Journal of the Institute of Actuaries*, **120**, 171-183.

Appendices

```
###Over-dispersed Poisson model
#For code the following book was used:,
#Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M. (2008).
#Modern Actuarial Risk Theory: Using R. Springer, Berlin, Heidelberg.
#p.281-289


#Data, Incremental
Cij <- c(4734994,1885305,281240,504341,524449,365049,100761,32449,3697,56901,
            4344093,1783774,243849,339985,49482,178961,508272,78125,1022,
            5288867,1795855,303246,351320,316038,33501,88774,31102,
            5357617,2548383,336749,403501,348378,236017,12982,
          5737732,2574724,971320,280140,226212,152127,
          5635064,2758392,241734,268113,429503,
          6629504,3045252,356119,200420,
          6824829,2669579,166400,
          8116439,3428535,
          10660074)

n <- length(Cij);
width <- trunc(sqrt(2*n))
m <- laius*(laius-1)/2          ## 45
p=2*laius-1                     ## number of parameters
i <- rep(1:laius,laius:1)
i <- as.factor(i)
j <- sequence(laius:1)
j <- as.factor(j)

###Estimate the upper triangle
CL <- glm(Cij~as.factor(i)+as.factor(j),family=quasipoisson)
###Find phi
Prs.resid <- (Cij-fitted(CL))/sqrt(fitted(CL))
#plot(Prs.resid)
phi.P <- sum(Prs.resid**2)/(n-p)
phi.P
#95229.07

###Estimate the full triangle
Cij.1 <- xtabs(Cij~i+j) ##full square matrix
ii <- row(Cij.1)
jj <- col(Cij.1)
Cij.1 <- as.vector(Cij.1)
future <- as.numeric(ii+jj-1 > laius) #future value is denoted with 1
ii <- as.factor(ii)
jj <- as.factor(jj)
#Fit the model
full.CL <- glm(Cij.1~ii+jj, fam=quasipoisson, wei=1-future)
```

```
sig <- vcov(full.CL) # A matrix of the estimated covariances between the parameter estimate
X <- model.matrix(full.CL)
cov.eta <- X%*%sig%*%t(X)
mu.hat <- fitted(full.CL)*future #fitted values
pe2 <- phi.P*sum(mu.hat) + t(mu.hat) %*% cov.eta %*% mu.hat
cat("Total reserve=",sum(mu.hat), "predition error=", sqrt(pe2),"\n")
#Total reserve= 13405108 predition error= 1985629

#ratio of process variance and estimation variance
phi.P*sum(mu.hat) / t(mu.hat) %*% cov.eta %*% mu.hat
# 0.4787986
```

```
##R-Code for Gamma model.
##For code the following book is used:
#Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M. (2008).
#Modern Actuarial Risk Theory: Using R. Springer, Berlin, Heidelberg. p. 282-289


#Data, Incremental
Cij <- c(4734994,1885305,281240,504341,524449,365049,100761,32449,3697,56901,
            4344093,1783774,243849,339985,49482,178961,508272,78125,1022,
            5288867,1795855,303246,351320,316038,33501,88774,31102,
            5357617,2548383,336749,403501,348378,236017,12982,
          5737732,2574724,971320,280140,226212,152127,
          5635064,2758392,241734,268113,429503,
          6629504,3045252,356119,200420,
          6824829,2669579,166400,
          8116439,3428535,
          10660074)

n <- length(Cij);
laius <- trunc(sqrt(2*n))
m <- laius*(laius-1)/2
p=2*laius-1
i <- rep(1:laius,laius:1)
i <- as.factor(i)
j <- sequence(laius:1)
j <- as.factor(j)

CL <- glm(Cij~as.factor(i)+as.factor(j),family=Gamma(link="log"))
summary(CL)
Prs.resid <- (Cij-fitted(CL))/fitted(CL)
phi.P <- sum(Prs.resid**2)/(n-p)
phi.P
#[1] 0.3217705

##Fill the lower triangle with initial values 1, to estimate the full triangle
Cij.1 <- t(matrix(c(4734994,1885305,281240,504341,524449,365049,100761,32449,3697,56901,
            4344093,1783774,243849,339985,49482,178961,508272,78125,1022,1,
            5288867,1795855,303246,351320,316038,33501,88774,31102,1,1,
```

```
            5357617,2548383,336749,403501,348378,236017,12982,1,1,1,
          5737732,2574724,971320,280140,226212,152127,1,1,1,1,
          5635064,2758392,241734,268113,429503,1,1,1,1,1,
          6629504,3045252,356119,200420,1,1,1,1,1,1,
          6824829,2669579,166400,1,1,1,1,1,1,1,
          8116439,3428535,1,1,1,1,1,1,1,1,
          10660074,1,1,1,1,1,1,1,1,1),
                nc=10, dimnames=list(dev=1:10, origin=1:10)))

ii <- row(Cij.1)
jj <- col(Cij.1)
Cij.1 <- as.vector(Cij.1)
future <- as.numeric(ii+jj-1 > laius)
ii <- as.factor(ii)
jj <- as.factor(jj)
full.CL <- glm(Cij.1~ii+jj, family=Gamma(link="log"), wei=1-future)
sig <- vcov(full.CL)
X <- model.matrix(full.CL)
cov.eta <- X%*%sig%*%t(X)
mu.hat <- fitted(full.CL)*future
pe2 <- phi.P*sum((mu.hat)**2) + t(mu.hat) %*% cov.eta %*% mu.hat
cat("Total reserve=",sum(mu.hat), "predition error=", sqrt(pe2),"\n")
#Total reserve= 12142220 predition error= 5411186
#ratio of process and estimation variance
phi.P*sum((mu.hat)**2) / t(mu.hat) %*% cov.eta %*% mu.hat
#0.3430256
```

```
##The origin code is from:
## http://lamages.blogspot.com/2013/01/reserving-based-on-log-incremental.html

##Log-normal model
##Data is given in a matrix, where future values are replaced with "NA"-s
tri <- t(matrix(c(4734994,1885305,281240,504341,524449,365049,100761,32449,3697,56901,
            4344093,1783774,243849,339985,49482,178961,508272,78125,1022,NA,
            5288867,1795855,303246,351320,316038,33501,88774,31102,NA,NA,
            5357617,2548383,336749,403501,348378,236017,12982,NA,NA,NA,
          5737732,2574724,971320,280140,226212,152127,NA,NA,NA,NA,
          5635064,2758392,241734,268113,429503,NA,NA,NA,NA,NA,
          6629504,3045252,356119,200420,NA,NA,NA,NA,NA,NA,
          6824829,2669579,166400,NA,NA,NA,NA,NA,NA,NA,
          8116439,3428535,NA,NA,NA,NA,NA,NA,NA,NA,
          10660074,NA,NA,NA,NA,NA,NA,NA,NA,NA),
            nc=10, dimnames=list(dev=1:10, origin=1:10)))

m <- dim(tri)[1]
n <- dim(tri)[2]
##Set data in form of a data frame
dat <- data.frame(
  origin=rep((1:m), n),
  dev=rep((1:n), each=m),
```

```
  value=as.vector(tri))
dat <- dat[order(dat$origin),]
dat <- with(dat, data.frame(origin, dev, cal=origin+dev,
                            value, logvalue=log(value),
                            originf=factor(origin),
                            devf=as.factor(dev),
                            calf=as.factor(origin+dev)))

rownames(dat) <- with(dat, paste(origin, dev, sep="-"))
dat <- dat[order(dat$origin),]

#Fit the model according to Log-Normal model
Fit <- lm(logvalue ~ originf + devf + 0, data=dat)
summary(Fit)

#Model design matrix
dm <- model.matrix(formula(Fit), dat=model.frame(Fit))

#Future design matrix
fdm <- model.matrix( ~ originf + devf + 0, data=dat[is.na(dat$value),])
#Calculate the variance-covariance matrix:
varcovar <- fdm %*% vcov(Fit) %*% t(fdm)
round(varcovar,4)

sigma <- summary(Fit)$sigma
sigma
# 0.6798913

Var <- varcovar + sigma^2
VarY <- Var[row(Var)==col(Var)]
Y <- fdm %*% coef(Fit)
P <- exp(Y + VarY/2)
VarP <- exp(2*Y + VarY)*(exp(VarY)-1) #Variance of lognormal distribution
seP <- sqrt(VarP)
i <- fdm %*% c((1:m)-1, rep(0, (n-1)))
j <- fdm %*% c(rep(0, (m-1)), (1:n)-1)
Results <- data.frame(i,j, Y, VarY, P, VarP, seP)
Results

#Fill the lower triangle with estimates
lower.tri <- xtabs(P ~ i+j, dat=Results)
lower.tri

#The Full Triangle
full.tri<- tri
full.tri[row(tri) > (nrow(tri) + 1 - col(tri))] <-
  lower.tri[row(lower.tri) > (nrow(lower.tri) - col(lower.tri))]


```

```
# Co-variance between the predictions
CoVar <-  sweep(sweep((exp(varcovar)-1), 1, P, "*"), 2, P, "*")
#Set the values on the diagonal to zero to use
#the variances calculated earlier (VarP),
#which includes the model variance sigma^2 as well.
CoVar[col(CoVar)==row(CoVar)] <- 0
round(CoVar,0)

#Find the overall variance
TotalVar <- sum(CoVar) + sum(VarP)
##The ratio of process and estimation variance
sum(CoVar)/sum(VarP)
# 0.4913061
#standard error
se <- sqrt(TotalVar)
se
# [1] 12851133
TotalReserve <- sum(lower.tri)
# [1] 17430668
cat("Total reserve=",TotalReserve, "standard error=", se,"\n")


##Finding the estimates for Mack model, the ChainLadder package is used.
library(ChainLadder)
M <- MackChainLadder(incr2cum(tri), est.sigma="Mack")
M$FullTriangle
M
```

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina Liivika Tee

sünnikuupäev: 29.04.1989

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Stochastic Reserving Methods in Non-Life Insurance,

mille juhendaja on Meelis Käärik,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **20.05.2013**